DOCUMENT RESUME

ED 442 863 TM 031 283

AUTHOR Osborne, Jason W.

TITLE Testing Stereotype Threat: Does Anxiety Explain Race and Sex Differences in Achievement?

PUB DATE 2000-04-00

NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.

DESCRIPTORS *Anxiety; Females; *High School Seniors; High Schools; *Low Achievement; Males; Mathematics; Minority Groups; Performance Factors; *Racial Differences; *Sex Differences; *Stereotypes

IDENTIFIERS High School and Beyond (NCES)

ABSTRACT

The stereotype threat theory of C. Steele (1992, 1997) attempts to explain the underperformance of minority students in academic domains and of women in mathematics. Steele asserts that situational self-relevance of negative group stereotypes in testing situations increases the anxiety these students experience, and that these differential anxiety levels explain performance differences. Research shows that manipulation of stereotype threat can affect academic performance. However, there has been little research testing whether anxiety does at least partially explain the relationship between race and achievement. The goal of this study was to examine whether anxiety will explain racial differences in academic performance and gender differences in mathematics performance in the context of a nationally representative sample of high school seniors. Data were drawn from the senior cohort data file of the High School and Beyond study, a cohort that initially consisted of 28,240 seniors from 1,015 schools. Partial mediation was observed, with anxiety explaining significant portions of the racial differences in academic performance. Anxiety also partially explained sex differences in mathematics achievement, although the effect sizes were very small. These results provide general support for Steele's stereotype threat hypothesis. (Contains 3 tables and 54 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 442 863

J. Osborne

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Testing Stereotype Threat: Does Anxiety Explain Race and Sex Differences in Achievement?

Jason W. Osborne

TM031283

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in New Orleans
April 2000

Testing Stereotype Threat: Does Anxiety Explain Race and Sex Differences in Achievement?

Jason W. Osborne
University of Oklahoma

ABSTRACT

Steele (1992, 1997) stereotype threat theory attempts to explain underperformance of minority students in academic domains and of women in mathematics. Steele argues that situational self-relevance of negative group stereotypes in testing situations increases the anxiety these students experience, and that these differential anxiety levels explain performance differences. Research shows that manipulation of stereotype threat can affect academic performance. However, there has been little research testing whether anxiety does at least partially explain the relationship between race and achievement. The goal of this study was to examine whether anxiety will explain racial differences in academic performance and gender differences in math performance in the context of a nationally representative sample of high school seniors. Partial mediation was observed, with anxiety explaining significant portions of the racial differences in academic performance. Anxiety also partially explained sex differences in math achievement, although the effect sizes were very small. These results provide general support for Steele's Stereotype Threat hypothesis.

Educational and Psychological research has repeatedly shown that students from disadvantaged minority groups tend to underperform in the academic realm relative to White or Asian students. For example, African-American students tend to receive lower grades in school (e.g., Demo & Parker, 1987; Simmons, Brown, Bush, & Blyth, 1978); score lower on standardized tests of intellectual ability (e.g., Bachman, 1970; Herring, 1989; Reyes & Stanic, 1988; Simmons et al., 1978), drop out at higher rates (e.g., American Council on Education, 1990; Steele, 1992), and graduate from college with substantially lower grades than White students (e.g., Nettles, 1988). There are similar findings in the literature for Latino students (e.g., Bruschi & Anderson, 1994; Ogbu, 1978; Whitworth & Barrientos, 1990) and Native American students (e.g., Ogbu, 1978).

These performance gaps are attributable to a host of factors, including socioecomonic status, academic preparation, educational opportunities, and so on. However, as authors such as Steele (1997) have pointed out, even when background factors (such as academic preparation) are held constant, subsequent achievement is lower for minority students than White or Asian students (Jensen, 1980; Ramist, Lewis, & McCamley-Jenkins, 1994). Further, this gap is not static, nor is it present at the beginning of schooling. The gap between White and minority students widens by as much as two grade levels by sixth grade (e.g., Alexander & Entwhistle, 1988; Valencia, 1991, 1997). In sum, there is convincing evidence that students from disadvantaged minority groups experience poorer outcomes at every level, even given equal preparation, than that of White and Asian students. Further, similar results are found for girls and women in "traditionally-male" domains, such as mathematics, computers, and engineering (e.g., Hyde, Fennema, & Lamon, 1990; Steele, 1997).

Claude Steele (1992, 1997) has argued that this underperformance of disadvantaged groups (racial minorities, girls and women in traditionally-male domains) is due to "stereotype threat." Steele argued that schooling and the school environment is aversive to members of these groups long before the achievement gap manifests because of negative stereotypes concerning the intellectual ability of group members. Specifically, Steele argued that while most students experience some anxiety over being negatively evaluated, students who belong to groups with a negative intellectual stereotype not only risk personal embarrassment and failure but also risk confirming the negative group stereotype. This, he argued, leads to increased anxiety, which depresses performance at every level of preparation in a number of different ways, as will be discussed below.

Note that Steele is presenting this as an explanation for the results of routinely-administered tests of intellectual or academic ability.[1] Steele (1997, p. 617) argued that stereotype threat is cued by the "mere recognition that a negative group stereotype could apply to oneself in a given situation." The negative group stereotype need not be believed, but merely be known. Thus, African-American or female students need not feel that they are, as a group, inferior to other students, but merely know that there is a stereotype present in the population to that effect. There is evidence that most members reared in a culture know of the stereotypes that are prevalent in that culture, regardless of whether they are believed or even consciously accessed (e.g., Devine, 1989). Thus, the argument is plausible.

Evidence of Stereotype Threat

There is mounting evidence supporting Steele's assertions. Psychological theory and research supports the basic assumption that awareness of a negative stereotype increases situational anxiety and evaluation apprehension as the domain becomes more self-relevant (e.g., Goffman, 1963, Howard & Hammond, 1985; Steele & Aronson, 1995; S. Steele, 1990). Further, there is evidence showing that increased anxiety or arousal can inhibit performance when that task at hand is complicated and/or not

BEST COPY AVAILABLE

well-learned (e.g., Geen, 1991; Hunt & Hillery, 1973; Michaels, Blommel, Brocato, Linkous, & Rowe, 1982; Sarason, 1972; Wigfield & Eccles, 1989) as is often the case with academic tasks. The racial performance gaps are therefore explained through the cognitive effects of increasing anxiety, which include decreasing cognitive capacity, reticence to respond, attentional deficits, and distracting thoughts, to name a few. Importantly, Steele argues that this is a general social-cognitive reaction to a situation, and that any group in this situation would react similarly (e.g., women in mathematics classes, other disadvantaged minority groups in school, etc.).

Anxiety and race. There is empirical support for the assertion that disadvantaged minority students do in fact have increased anxiety in while performing in academic arenas. For example, Payne (1984) reported that African-American students not only scored higher on a measure of test anxiety, but also demonstrated more behavioral indices of anxiety (answer changing). Silverman, La Greca, & Wasserstein (1995) reported higher anxiety in African-American elementary school children than White or Hispanic children. Additionally, Clawson, Firment, and Trower (1981) found higher state and trait levels of anxiety among African-American seventh-and ninth-graders when compared to White students, and reported that higher-anxiety students (regardless of race) scored significantly lower than less anxious students on achievement tests. Unfortunately, the authors failed to test whether anxiety mediated the relationship between race and academic outcomes, one of the central assertions of Steele's theory.

Stereotype threat and test performance. The implication of stereotype threat is that removal of the threat would close the racial performance gap (assuming student backgrounds, especially academic preparation, are largely equivalent). The results of experimental investigations into this phenomenon have been encouraging. For example, Steele and Aronson (1995) found that when an extremely difficult test of verbal abilities was described to Stanford undergraduates as diagnostic of ability (therefore triggering stereotype threat), African-Americans scored significantly lower than Whites. However, when the test was presented as a laboratory problem-solving task (and thus non-diagnostic of ability and not triggering stereotype threat), the performance of African-Americans was not substantially different than that of White students, and was significantly better than African-American students in the threat condition. Subsequent experiments reported by Steele and Aronson (1995) showed that in the stereotype-threat condition, racial stereotypes were activated to a greater degree than in the non-stereotype threat condition, and that when given a chance to self-identify one's race, all students did so except for the African-American students in the stereotype threat condition, where only one-fourth did. This last finding may represent a desire to avoid the anxiety-provoking stereotype (Aronson, Quinn, & Spencer, 1998). Combined, these studies reported in Steele & Aronson (1995) support the notion of stereotype threat.

Although much of the discussion and evidence pertains to African-American and White students, this theory of stereotype threat is more general than that. As any group in similar circumstances should experience the same results, other studies (summarized in Aronson, Quinn, & Spencer, 1998) have examined this hypothesis in other populations where there is a stereotype of intellectual inferiority, such as with Latino students (Aronson & Salinas, 1997), and girls and women in traditionally-male domains, such as math and science (Shih, Pittinsky, & Ambady, 1999; Spencer, Steele, & Quinn, 1999). Additionally, Spencer, Steele, & Quinn (1999) reported that varying that applicability of the stereotype (in this case, gender) influenced the results in the same way: when a stereotype was perceived to be unrelated to a task results were more equal (women scoring as high as men) than when the applicability of the stereotype was not undermined (see also Broadnax, Crocker, & Spencer, 1997 for similar results for African-American and White students). In a powerful demonstration of the universality of this phenomenon by Aronson, Lustina, Good, Keough, Steele, and Brown (1999), highly-math proficient White males were induced into feeling stereotype threat while taking a math test by being told that the purpose of the experiment was to understand why Whites did so poorly on the (particular) exam compared to Asian students. As expected, White males in the stereotype threat condition performed significantly worse than White males in a no-threat condition. In a similar study of extremely math-talented Asian-American female undergraduates, Shih, Pittinsky, and Ambady (1999) demonstrated that Asian-American females' performance on a math achievement test was enhanced when their Asian identity (and hence the positive Asian and math stereotype) was made most salient, and degraded when their female identity (and hence the negative female and math stereotype) was made most salient, relative to a control group of Asian-American females. Importantly, in this study, groups did not differ on observed motivation, perceptions of test performance, and were not aware that a particular target identity was being made more salient.

These studies support the following assertions: (a) stereotype threat is situationally-specific and not a trait of a group, (b) stereotype threat is a phenomenon that any group can fall victim to if they are in a situation where there is a negative group stereotype concerning their performance in that domain, the domain is self-relevant, and the stereotype is known, (c) groups in these conditions seek to escape the stereotype threat by any means possible (e.g., not acknowledging membership in the stigmatized group), (d) belief of the stereotype is not a necessary condition, and (e) that reducing stereotype threat improves the performance of the stigmatized group to the point where performance is not substantially different from that of non-stigmatized groups. This last point, repeatedly demonstrated in the studies mentioned above, is the main reason why this theory is the focus of much interest and attention.

Mediators

There are several possible explanations for the observed results. First, Steele and colleagues (e.g., Aronson, Quinn, & Spencer, 1998; Steele, 1997) argue that anxiety explains (mediates) the observed experimental effects. However, there are other possible explanations. Most prominent is the assertion that it could simply be that when a student is in a situation for

BEST COPY AVAILABLE

which there is a negative group stereotype, that student has lower expectations or efficacy. A study by Spencer, Steele, and Quinn (1999) tested three possible mediators in a sample of women and men taking a difficult version of a standardized math test, with the same types of stereotype threat/no threat manipulations as described above: state anxiety, evaluation apprehension, and self-efficacy. Results showed that, of the three only anxiety was found to be a partial mediator of the relationship. Further, the results from Shih et al. (1999) help to rule significant differences in motivation, perceived performance, liking for the test, assessment of test difficulty, or assessment of personal ability. Thus, anxiety appears to be a good explanation for the effect. Of course, other potential mediators, such as distractibility (which could be related to anxiety) and distrust of the system have also been discussed, and should be pursued as well.

Summary and Goals of the Present Study

In sum, there appears to be growing evidence for stereotype threat as an explanation for the relatively poor academic outcomes that disadvantaged minority students receive, and some evidence supporting the role of anxiety as the mechanism (or a major aspect of the mechanism) through which stereotype threat affects academic outcomes. More psychologists and educators are becoming interested in this phenomenon. However, several notes of caution need to be sounded.

First, the predominant population studied thus far has been high achieving college undergraduates. While this makes for powerful demonstrations of the theory, it does not address stereotype threat in the mainstream population.

Second, while there is convincing evidence that stereotype threat manipulations, and perhaps stereotype threat in general, can affect academic outcomes, there is little empirical evidence concerning the mediating mechanism. For example, Spencer, Steele, and Quinn (1999) report that anxiety mediates the effects of stereotype threat in women taking a mathematics test, but another unpublished study (Broadnax, Crocker, & Spencer, 1997, reported in Aronson, Quinn, and Spencer, 1998) reported that for African-Americans, evaluation apprehension (and not anxiety, which was measured separately) mediated the effects of stereotype threat. Both of these studies have been from convenience samples of post-secondary students, neither offering compelling evidence regarding the primary hypothesized link in Steele's theory between the stereotype and the outcome: anxiety. Thus, stereotype threat is still an effect without a clearly explicated mechanism.

Third, while Steele's theory should apply to any group for which there is a stereotype of intellectual inferiority, most of the research focuses either on African-American students or women in mathematics domains. However, Latino and Native American students should show similar effects (as they also tend to be stigmatized), yet have received little attention.

The goal of the present study is to examine Steele's stereotype threat theory in a nationally-representative sample of secondary school students in two ways. The first part of this study will examine whether anxiety explains at least part of the racial differences in achievement test scores between Whites and students of color (African-Americans, Latinos, and Native Americans). The second part of the study will examine whether anxiety explains at least part of the gender differences in math performance. This would be the first study to examine this mediational hypothesis in a sample not drawn from elite undergraduates or convenience samples, which would provide more generalizability to the population of students as a whole.

METHOD

Overview

Data for this investigation were drawn from the senior cohort data file from the High School and Beyond (HS&B) study. This study was initiated in 1980 by the National Center for Educational Statistics (NCES). The senior cohort initially consisted of 28,240 seniors from 1,015 schools in the United States.

Variables

Achievement test scores. During the course of the data collection session for this study, students completed achievement tests covering six topics: vocabulary, reading, mathematics, three-dimensional visual spatial ability, perceptual comparisons, and mnemonic memory skills. Because stereotype threat requires students to believe that the task is at least a modestly valid measure of the domain being assessed, the three sections that students were most likely to have had experience with were selected for inclusion in this study: vocabulary, reading, and mathematics.[2]

The vocabulary test consisted of twenty-seven synonym-format items. The reading test consisted of twenty items relating to several short (100-200 word) passages. These items focused primarily on comprehension, although some questions relating to analysis and interpretation were included. The mathematics test consisted of thirty-three items, in a format where students were asked to indicate which of two quantities is greater. All sections were timed.

These tests were administered at the end of the same session where the other data were gathered. Standardized scores on these three tests were well correlated (rs ranged from .58 to .60), and had an estimated internal reliability (Cronbach's alpha) of $\alpha = .81$. Thus, a composite achievement test score was created by averaging the three standardized scores.

Anxiety. Immediately after the tests, subjects completed a short battery that asked how they felt while taking the tests. All were measured on a dichotomous yes (1) and no (0) scale, with the item stem "How did you feel while you were taking the tests?" and included the following items: tense, under pressure, under strain, nervous/jittery, uneasy, calm, afraid of not doing well, and uncomfortable. Factor analyses revealed a clear single factor (eigenvalue=3.89, factor loadings ranging from .41 to .70 in magnitude, no other factor with an eigenvalue greater than 1.0). All items were scored so that increasing scores indicated increasing anxiety. Cronbach's alpha for this scale was $\alpha = .79$. A composite anxiety score was created by summing these variables. The composite variable had significant positive skew, ranging from 0 to 8, with less than 1.1% of subjects scoring over a 6. To reduce skew, all scores over 6 were recoded to 6.

BEST COPY AVAILABLE

Additionally, a quadratic anxiety term was created by squaring the variable.

Academic Preparedness. Prior academic preparedness was assessed by student reports of average grades. Students were asked to rate their grades in high school on a scale from mostly As (1) to mostly Fs (8).[3] Prior preparedness was found to correlate with achievement test scores (r = .53) in the general sample.

Demographic variables. Sex, race, and ethnic background variables were gathered during the same data collection session as the other variables. Five racial groups were formed from the combination of race and ethnic background variables. Students who indicated that their race was White, and that their background was of European or North American descent were considered White (e.g., American, Canadian, German, French, Polish, etc.; N=12,557). Students indicating that their race was Black and that their background was Afro-American (not Caribbean or West Indian) or American descent were considered African-American (N=1,846). Students indicating that they were of Mexican/Chicano, Cuban, Puerto-Rican, or Latin-American descent were considered Latino (N=1,047). Finally, students indicating that their race is American Indian, or Alaskan were identified as Native American (N=111). Many cases were eliminated from the data set as a result of this grouping, as the goal was to form the most homogenous groups possible. For example, students indicating that they are Black but listing ethnicity as Mexican or Irish, or students listing their race as White but their ethnicity as Afro-American, were excluded from analyses due to this ambiguity. Additionally, only cases with complete data on all relevant variables were included.

Analytical Plan

The first hypothesis to be tested is that racial differences in academic outcomes (specifically, achievement test scores) are at least partially explained by differing levels of anxiety while engaged in academic evaluation or while in the academic environment. According to Steele's theory of stereotype threat, students belonging to disadvantaged minority groups should be more anxious while taking tests of academic or intellectual ability, as failure to perform well could confirm negative group stereotypes.

The second hypothesis to be tested is that gender differences in math performance (math achievement test scores) are at least partially mediated by differential levels of anxiety. Along similar lines, Steele has argued that girls perform poorly relative to boys because of a stereotype that girls are inferior to boys in math performance. The hypothesis to be tested in both cases that anxiety will at least partially mediate the relationship between race or gender and academic outcomes.

According to Baron and Kenny (1986), four conditions must be satisfied in order to claim mediation: (a) variations in the independent variable account for significant variance in the mediator, (b) variations in the mediator account for significant variance in the dependent variable, (c) the independent variable accounts for significant variance in the dependent variable, and (d) when the mediator is controlled for, the relationship between the independent variable and the dependent variable is significantly decreased. These four conditions will be tested separately for each of the two major hypotheses (race, sex).[4]

RESULTS

Does Anxiety Explain Racial Differences in Academic Achievement?

During initial analyses, sex did not to interact with race in predicting achievement test outcomes (i.e., the effects of race were not modified by sex), and was omitted from these analyses. For these analyses (examining group differences), group size differences are undesirable. Thus, three separate analyses were conducted, one for each of the White vs. non-white contrasts: White vs. African-American, White vs. Latino, and White vs. Native American. The advantage of this procedure is that groups can be randomly sampled to be exactly identical in size, and it does not mask differences between the three groups of color, whereas a single analysis with a single four-level race variable might. Thus, for the White/African-American analyses, the White group was randomly sampled so each group contained 1,846 cases. For the White/Latino analyses, the White group was randomly sampled so that each group contained 1,047 cases. For the White/Native American analyses, the White group was randomly sampled so that each group contained 111 cases. For each sampling, full replacement was used in the White group.

Relationship of race to achievement. Three univariate ANOVAs were performed, one for each of the White/non-White contrasts. The results of these analyses, which are presented in Table 1, indicate large and highly significant differences in achievement test performance as a function of race, with race accounting for between 9% and 23% of the variance in achievement test scores. As expected, the White students performed significantly higher than any of the non-White groups.

Insert Table 1 about here

Relationship of race to anxiety. Three univariate ANOVAs were performed, one for each of the White/non-White contrasts. The results of these analyses, which are presented in Table 2, indicate highly significant differences in anxiety relating to test taking as a function of race, with race accounting for between 4% and 5% of the variance in anxiety. As expected, White students had lower anxiety scores than the non-White students.

Insert Table 2 about here

Relationship of anxiety to achievement test scores. In order to test whether anxiety relates to achievement test, a multiple regression was computed predicting achievement from anxiety and a quadratic anxiety score. The multiple correlation between the two anxiety variables and achievement was R = -0.31 (p < .0001).

Test of mediation. All three previous conditions for mediation were met for all three contrasts between racial groups. One final analysis examined whether the introduction of the mediators into an analysis would significantly attenuate the relationship between race and achievement test scores. For each

BEST COPY AVAILABLE

racial contrast, two multiple regression analyses were performed. The first included only the race variable predicting achievement test scores. The second was a blockwise multiple regression analysis, entering the anxiety and a squared anxiety term (as the anxiety and performance literature has reported curvilinear relationships between these two variables) on the first step, and the dummy-coded race variable on the second step. To test whether the relationship between race and achievement was significantly reduced, the difference between the two regression coefficients for race were tested for significance according to procedures outlined in Clogg, Petkova, and Haritou (1995a, 1995b) for testing nested regression coefficients.

Insert Table 3 about here

The results of these analyses, presented in Table 3, indicate that, while anxiety does not completely mediate the relationship between race and achievement, the race effects on achievement were partially mediated in two of the three analyses (White vs. African-American, White vs. Latino). In both cases, race accounted for between 21% and 23% of the variance in achievement test scores, and in both cases anxiety (and its quadratic component) explained between 38.8% and 41.4% of the effect of race.

Because anxiety was measured after completing the last item of the test, there is a question of reciprocity. Specifically, it is possible that differences in the test outcome caused the differences in anxiety, rather than differences in anxiety causing differences in test outcomes, as Steele (e.g., 1997) posits. To attempt to address this issue two analyses were performed: one replicating the previous analyses controlling for academic preparation, and one testing whether there are significant differences in anxiety once test scores are covaried (i.e., to test whether differences in anxiety are wholly attributable to differences in test scores or not). Given the limitations of the data set, this is as rigorous a test of reciprocity as is possible.

Analyses covarying prior academic preparation. As seen in Table 3, after covarying prior academic preparedness, anxiety significantly mediated the relationship between race and achievement for the same two analyses (White vs. African-American, White vs. Latino). In these two analyses, race accounted for between 10% and 15% of the variance in achievement test scores, once prior preparation was covaried. Of this effect, anxiety accounted for between 28% and 31% of the effect. Thus, covarying academic preparation reduced both the magnitude of the racial differences as well as the magnitude of the mediation effect. However, both are still large enough to be of theoretical interest.

Analyses predicting anxiety controlling for achievement test scores. In order to test whether there were significant racial differences in anxiety once achievement was controlled, three separate ANOVAs were calculated, each contrasting Whites with either African-Americans, Latinos, or Native Americans. Again, the first two contrasts (White vs. African-American, White vs. Latino) showed significant differences in anxiety once achievement was controlled (all $p < .0001$). The White vs. Native American analysis did not show significant differences once achievement was covaried.

Summary. All requirements for partial mediation, as put forth by Baron and Kenny (1986) were met. Specifically, the independent variable was related to both the dependent variable and the mediator variable, the mediator variable was related to the dependent variable, and when the mediator were controlled for, the effect of the independent variable was reduced significantly and substantially, but not completely. Further, even after controlling for prior preparation, anxiety was a significant mediator. Further, one other test showed that there were still anxiety differences once achievement was covaried for two of the three contrasts, lending some further support to the assertion that anxiety is mediating test performance, rather than vice-versa. However, the issue is still far from resolved here.

Also of interest is the fact that anxiety did not appear to explain the achievement differences between White and Native-American students. While these two groups have not been addressed in the literature, there was good theoretical reason to expect this comparison to follow the same pattern as that of the other two disadvantaged minority groups.

Does Anxiety Explain Sex Differences in Math Achievement?

For these analyses, all students with valid data were retained (10,320 males, 11,510 females). Females were randomly sampled to equalize cell sizes at 10,320. Analyses similar to those performed above were repeated here, exploring whether sex differences in math performance are mediated by differences in anxiety, as Steele (1992, 1997) asserted.

Relationship of sex to achievement. An ANOVA was performed, with sex as the independent variable and math achievement scores as the dependent variable. The anticipated effect was observed, with males (mean = 0.14) outscoring the females (mean = - 0.14, $F=520.30$, $p < .0001$). While highly significant, the effect size is relatively small, with an eta-squared of .03.

Relationship of sex to anxiety. Another ANOVA, examining sex differences in anxiety, was performed. As expected, males showed lower levels of anxiety (mean=1.19) than females (1.43, $F = 97.20$, $p < .0001$, eta-squared less than .01).

Relationship of anxiety to achievement test scores. A multiple regression was computer predicting math achievement scores from anxiety and its quadratic component. The multiple correlation was $R = -.26$, $p < .0001$).

Test of mediation. All three previous conditions for mediation were met. One final analysis examined whether the introduction of the mediators into an analysis would significantly attenuate the relationship between sex and math achievement test scores. In order to test this, two multiple regression analyses were performed. The first included only sex predicting math achievement test scores. The second included anxiety and its quadratic component on the first step, and sex on the second step.

The results of these analyses indicate that sex accounted for a significant but small amount of variance in math achievement ($\beta = -.16$, $p < .0001$), and that when anxiety was covaried, the

BEST COPY AVAILABLE

beta was reduced to $\beta = -.14$, $p < .0001$. While very small, this reduction was significant ($z = 6.52$, $p < .0001$). When prior academic preparedness was controlled for, covarying anxiety reduced the beta from -0.24 to -0.22, which was also significant ($z = 4.59$, $p < .0001$). Finally, an analysis examining sex differences in anxiety once achievement was covaried revealed highly significant ($p < .0001$) differences, again lending support to the argument that anxiety is mediating achievement test performance, not the reverse.

Summary. All requirements for partial mediation, as put forth by Baron and Kenny (1986) were met. While this analysis showed significant differences in regression coefficients once anxiety was covaried, the magnitude of the differences were very small. The effect sizes for sex were very small, and as such these analyses are fairly unimportant in terms of theory.

DISCUSSION

The goal of this study was to test Steele's theory of stereotype threat in a representative sample of high school students. The specific hypothesis tested was Steele's (e.g., 1997) assertion that anxiety would explain racial group-level differences in achievement test scores and gender differences in math achievement scores. The results of the analyses generally supported a partial mediation explanation. Anxiety (and its quadratic component) accounted for a significant portion of the achievement test differences between Whites and African-Americans, and between Whites and Latinos. Anxiety was not found to explain the differences between White and Native American students, which may have been attributable to either small sample sizes or fundamental differences in the mechanism explaining achievement differences with this group. As little attention has been given to Native Americans in the stereotype threat literature, this should be a topic for future research.

The two analyses showing significant mediation also showed that up to 23% of the variance in achievement test scores are attributable to racial differences. Further, up to 41% of these effects are explained by differing anxiety levels. Even after prior academic preparation was taken into account, up to 15% of the variance in achievement test scores are attributable to race, and up to 31% of this effect was accounted for by differences in anxiety. Finally, ancillary analyses showed that anxiety differences were not wholly attributable to test score differences.

The results for sex differences in math achievement were much less spectacular, in part due to the small sex differences in math achievement. While both analyses (covarying anxiety, and covarying anxiety and prior academic preparedness) produced significant differences in regression weights, it is largely due to the enormous sample size. While technically supporting Steele's assertions, the small effects produced in these analyses suggest that other research is needed to determine whether widely-noted sex differences on math and science-related tasks are explained by stereotype threat. It is possible that the math test present in this data set is not representative of math tests that show large gender differences cited in other research, for example.

Caveats

There are several caveats that need to be addressed with this study. First, because anxiety was measured after the test was completed, it is not entirely clear that anxiety can be said to have caused differences in test performance. Two sets of analyses attempted to clarify the causality relationship between anxiety and achievement, generally supporting the assertion that anxiety caused the differences, but as is always the case with correlational data, this is a tenuous assertion, and should not be taken too seriously until and unless experimental research supports this conclusion. While it is unclear to what extent students were able to perceive their own performance accurately (as no feedback was given to them regarding their performance), it is clear that other ways of testing this hypothesis should be employed in future investigations of this hypothesis (e.g., real-time measures of anxiety, such as psychophysiological measures).

Second, there is a question as to whether stereotype threat existed during this test administration. As stated above, in experimental research by Steele, Aronson, Spencer, and others, conditions under which stereotype threat are assumed to occur (there is as yet no direct measure of stereotype threat, merely data that allow inference of stereotype threat) have included: stereotypes that are clearly relevant to the task, high task difficulty, clear diagnosticity of the task, and high self-relevance of the task at hand, among other conditions. However, these are experimental conditions set up to provide large effect sizes when performance under these conditions are compared to other conditions. According to Steele (1997, 1998), the necessary conditions for stereotype threat include a situation where there is a negative group stereotype concerning their performance in a domain, self-relevance of the domain, and awareness of the negative group stereotype (note that belief in the stereotype is not necessary-- mere knowledge of the stereotype is sufficient), and belief that the task undertaken reveals ability in the domain. It is not a stretch to assert that these conditions prevail at most standardized testing sessions in this country, and in this testing session in particular.

The presence of students who did not view academics as self-relevant is an issue with the study, however. In particular, some of the minority students might have been disidentified with academics (i.e., students might view academics as not self-relevant, see Osborne, 1997; Steele, 1997) at the time of testing. If this were the case the mediation effect of anxiety would be underestimated, and the results would be biased toward the null hypothesis. Thus, it is possible and likely that in a sample of high-school students that are all strongly identified with academics the medication effect of anxiety would be stronger. Alternatively, as students knew they were taking these tests as part of research, it is possible that students viewed these tasks as less important than other, routine tests they might take, and therefore this research might have underestimated the effects reported within.

Third, in any analysis where there is a covariate, to the extent that the covariate was measured with less than perfect reliability, the complete effect will not be removed, again biasing the results toward the null. As the anxiety measure had less than ideal reliability, these analyses were replicated using structural equation modeling to overcome the measurement

BEST COPY AVAILABLE

limitations. The results from this analysis showed slightly larger effect sizes, and slightly larger mediation effects for anxiety (however, it was not possible to model the quadratic anxiety term). Additionally, all reductions in regression weights were highly significant, as the standard errors of the estimate were smaller. However, as these analyses did not provide any substantively different results or conclusions, and given the inability to model a curvilinear relationship with a latent variable, and the higher complexity of reporting SEM analyses, the OLS regression results were deemed simpler and equally informative.

Finally, while the goal of this study was to examine a sample of secondary school students (rather than elite college students), this study examines high school seniors. While this might seem but a little removed from the other samples cited in this article, there is one major advantage to this sample—external validity. Although developmentally this sample and the sample from other studies are similar, the students in the present study represent the broad spectrum of high school students in America, not just those students attending elite colleges and universities.

Summary and Directions for Future Research

This study, and others that have preceded it, indicate that stereotype threat is a viable and important theory. While other theories on racial differences in achievement have focused on relatively fixed and immaleable constructs, such as cognitive style (e.g., Shade, 1982), an oppositional group culture and peer pressure not to achieve (e.g., Fordham, 1988; Ogbu, 1992), a tendency for African-American students to avoid intellectual competition (e.g., Howeard & Hammond, 1985) and even genetic differences (e.g., Herrnstein & Murray, 1994), this theory focuses on the proximal environment while engaging in educational or academic tasks, which is more easily manipulated and intervened upon than genetics, cognitive style, and the like.

Based on this theory, Steele (1992, 1997) and others (e.g., Aronson, Quinn, & Spencer, 1998; Osborne, in press) have listed recommendations for improving the outcomes of students belonging to disadvantaged minority students. These recommendations include: inoculating students against stereotype threat by emphasizing the learning curve, and how far students have progressed prior to an evaluation, assisting students in revising their notions of intelligence from that of a fixed quantity to a more malleable trait, use of the Jigsaw Classroom technique, and emphasizing challenge and effort rather than talent or innate ability, to name a few (for a more thorough review of studies of interventions based on this theory, see Aronson, Quinn, & Spencer, 1998). Steele (1992, 1997) and Osborne (in press) have also discussed system-level changes that would help undermine the dynamics involved in stereotype threat. Among these changes are the implementation of a true multicultural curriculum, rather than what is often little more than brief studies of cultural difference (e.g., Grant & Sleeter, 1985). Along this vein, anything that the educational system, individual schools, individual teachers, community groups, parents, and peers can do to undermine the negative group stereotypes concerning intellectual ability will help improve the academic outcomes of members of those groups.

While the literature on stereotype threat is still growing and emerging, there is currently enough supporting evidence to be tantalizing and exciting. This phenomenon deserves more attention, particularly at earlier ages to understand at what developmental stages stereotype threat becomes an issue. For example, studies by Osborne (1995, 1997) and Hyde et al. (1990) identify the junior high years as the developmental period where stereotype threat, and it's ultimate consequence, academic disidentification (Major & Schmader, 1998; Steele, 1997), may appear. It is likely that interventions designed to thwart the effects of stereotype threat are probably most efficacious as early as possible in a child's academic career, but a lack of research leaves us with no information as to when the critical periods are for intervention.

Further, more research is needed with other groups besides African-Americans and women in math and science. Other disadvantaged minority students (e.g., Latino, Native American) suffer similar performance deficits in academics, and theory and empirical evidence suggests that they should benefit from similar interventions.

Additionally, as mentioned above, more rigorous tests of this anxiety mediation hypothesis should be undertaken, preferably using real-time psychophysiological measures of anxiety under well-controlled, experimental conditions where the problems present in these analyses can be eliminated or minimized.

Finally, stereotype threat is probably not a phenomenon specific to academics. Research should look at other arenas (e.g., sports competition, business and industry, and government service, for example) for this phenomenon.

BEST COPY AVAILABLE

REFERENCES

Alexander, K. L., & Entwhistle, D. R. (1988). Achievement in the first two years of school: Patterns and processes. Monographs of the Society for Research in Child Development, 53(2).

Allison, P. D. (1995). The impact of random predictors on comparisons of coefficients between models: Comment on Clogg, Petkova, and Haritou. American Journal of Sociology, 100, 1294-1305.

American Council on Education. (1990). Minorities in Higher Education. Washington, DC: Office of Minority Concerns.

Aronson, J., Lustina, M., Good, C., Keough, K., Steele, C. M., & Brown, J. L. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. Journal of Experimental Social Psychology, 35, 29-46.

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.) Prejudice: The Target's Perspective. New York: Academic Press.

Aronson, J., & Salinas, M. F. (1997). Stereotype threat, attributional ambiguity, and Latino underperformance. Unpublished manuscript, University of Texas.

Bachman, J. G. (1970). Youth in Transition: Volume 2. The impact of family bachground and intelligence on 10th-grade boys. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51, 1173-1182.

Broadnax, S., Crocker, J., 7 Spencer, S. (1997). African-Americans and academic performance: The medicational role of stereotype vulnerability. Unpublished manuscript: State University of New York at Buffalo.

Bruschi, B.A., & Anderson, B.T. (1994, February). Gender and ethnic differences in science achievement of nine-, thirteen-, and seventeen-year-old students. Paper presented at the Annual Meeting of the Eastern Educational Research Association, Sarasota, Florida.

Clawson, T. W., Firment, C. K., & Trower, T. L. (1981). Test anxiety: Another origin for racial bias in standardized testing. Measurement and Evaluation in Guidance, 13, 210-215.

Clogg, C. C., Petkova, E., & Haritou, A. (1995a). Statistical methods for comparing regression coefficients between models. American Journal of Sociology, 100, 1261-1293.

Clogg, C. C., Petkova, E., & Haritou, A. (1995b). Reply to Allison: More on comparing regression coefficients. American Journal of Sociology, 100, 1305-1312.

Demo, D.H. & Parker, K.D. (1987) Academic achievement and self-esteem among African-American and White college students. Journal of Social Psychology, 127, 4, 345-355.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5-18.

Fordham, S. (1988). Racelessness as a factor in Black students' school success: Pragmatic strategy or Pyrrhic victory? Harvard Educational review, 58, 54-84.

Geen, R. G. (1991). Social motivation. Annual Review of Psychology, 42, 377-399.

Goffman, E. (1963). Stigma: Notes on the management of a spoiled identity. New York: Touchstone.

Grant, C. A., & Sleeter, C. E. (1985). The literature on multicultural education: Review and analysis. Educational Review, 37, 97-118.

Herring, R. D. (1989). Minimum competency testing: Implications for ethnicity. Journal of Multicultural Counseling and development, 17, 98-104.

Herrnstein, R. J., & Murray,C. (1994) The bell curve: Intelligence and class structure in American life. New York: Free press.

Howard, J., & Hammond, R. (1985, September 9). Rumors of inferiority. New Republic, 72, 18-23.

Hunt, P. J., & Hillery, J. M., (1973). Social facilitation in a location setting: An examination of the effects over learning trials. Journal of Experimental Social Psychology, 30, 563-571.

Hyde, J.S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin, 107, 139-155.

Jensen, A. R. (1980). Bias in Mental Testing. New York: Free Press.

Major, B., & Schmader, T. (1998). Coping with stigma through psychological disengagement. In J. K. Swim & C. Stangor (Eds.) Prejudice: The Target's Perspective. New York: Academic Press.

Michaels, J. W., Blommel, J. M., Brocato, R. M., Linkous, R. A., & Rowe, J. S. (1982). Social facilitation and inhibition in a natural setting. Replications in Social Psychology, 2, 21-24.

National Center for Educational Statistics. (1984). High School and Beyond, 1980: A Longitudinal Survey of Students in the United States. U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Educational Statistics. (1992). National Education Longitudinal Study of 1988 First Follow-up: Student component data file user's manual. U.S. Department of Education, Office of Educational Research and Improvement.

Nettles, M. T. (1988). Toward undergraduate student equality in American higher education. New York: Greenwood.

Ogbu, J. U. (1978). Minority education and caste: The American system in cross-cultural perspective. New York, Academic Press.

Ogbu, J. U. (1992). Understanding cultural diversity and learning. Educational Researcher, 21, 5-14.

Osborne, J.W. (1995). Academics, self-esteem, and race: A look at the underlying assumptions of the disidentification hypothesis. Personality and Social Psychology Bulletin, 21(5), 449-455.

Osborne, J. W. (1997). Race and academic disidentification. Journal of Educational Psychology, 89, 728-735.

BEST COPY AVAILABLE

Osborne, J. W. (April, 1999). Historical trends in race and academic disidentification, 1972-1992. Paper presented at the national meeting of the American Education Research Association (AERA), Montreal, Quebec.

Osborne, J. W. (in press). Academic Disidentification: Unraveling Underachievement among Black Boys.

Payne, B. (1984). The relationship of test anxiety and answer changing behavior: An analysis by race and sex. Measurement and Evaluation in Guidance, 16, 205-210.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). Student group differences in predicting college grades: Sex language and ethnic groups (College Board Report No. 93-1, ETS No. 94.27). New York: College Entrance Examination Board.

Reyes, L. H., & Stanic, G. M. (1988). Race, sex, socioeconomic status, and mathematics. Journal for Research in Mathematics Education, 19, 26-43.

Sarason, I. G. (1972). Experimental approaches to test ansiety: Attention and the uses of information. In C. D. Spielberger (Ed.) Anxiety: Current trends in theory and research, 2. New York, Academic Press.

Shade, B. (1982). Afro-American cognitive style: A variable in school success. Review of Educational Research, 52, 219-244.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. Psychological Science, 10, 80-83.

Silverman, W. K., La Greca, A. M., & Wasserstein, S. (1995). What do children worry about? Worries and their relation to anxiety. Child Development, 66, 671-686.

Simmons, R. G., Brown, L., Bush, D. M., & Blyth, D. A. (1978). Self-esteem and achievement of Black and White adolescents. Social Problems, 26, 86-96.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. Journal of Experimental Social Psychology, 35, 4-28.

Steele, C. (1992, April). Race and the schooling of African-American Americans. The Atlantic Monthly, 68-78.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52, 613-629.

Steele, C. M. (1998). Stereotyping and its threat are real. American Psychologist, 53, 680-681.

Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. Journal of Personality and Social Psychology, 69(5), 797-811.

Steele, S. (1990). The content of our character. New York: St. Martin's Press.

Valencia, R. R. (1991). The plight of Chicano students: An overview of schooling conditions and outcomes. In R. R. Valencia (Ed.) Chicano school failure and success: Research and policy agendas for the 1990s (pp. 3-26). The Stanford Series on Education and Public Policy. London, England: Falmer Press.

Valencia, R. R. (1997). Latinos and education: An overview of sociodemographic characteristics and schooling conditions and outcomes. In M. Barrera-Yepes (ed.) Latino education issues: Conference proceedings. Princeton, NJ. Educational Testing Service.

Whitworth, R.H., & Barrientos, G.A. (1990). Comparison of Hispanic and Anglo graduate record examination scores and academic performance. Journal of Psychoeducational Assessment, 8(2), 128-132.

Wigfield, A., & Eccles, J. S. (1989). Test anxiety in elementary and secondary school students. Educational Psychologist, 24, 159-183.

AUTHOR NOTES:
Jason W. Osborne, Ph.D.,
Department of Educational Psychology,
University of Oklahoma
820 Van Vleet Oval
Norman, OK, 73019
(405) 325-1067
(405) 325-6655 (fax)
josborne@ou.edu

BEST COPY AVAILABLE

Table 1
Relationship between Race and Achievement Test Scores.

Contrast:	White	Non-White	F	Eta^2
White vs. African-American (N=1846/1846)	0.22	-0.58	1000.02***	.23
White vs. Latino (N=1047/1047)	0.26	-0.45	439.51***	.21
White vs. Native American (N=111/111)	0.16	-0.34	21.03***	.09

Notes: Achievement test scores were z-scored. Cell sizes are presented in parentheses. *** $p < .0001$.

Table 2
Relationship between Race and Anxiety.

Contrast:	White	Non-White	F	Eta^2
White vs. African-American (N=1846/1846)	1.10	1.81	144.78***	.04
White vs. Latino (N=1047/1047)	1.06	1.79	92.32***	.04
White vs. Native American (N=111/111)	1.02	1.84	11.88***	.05

Notes: Cell sizes are presented in parentheses. *** $p < .0001$.

Table 3
Test of mediation.

							Covarying Preparation					
Contrast	β for Race, Model 1	R^2 for Race. Model 1	β for Race, Model 2	ΔR^2 for Race, Model 2	Z for difference	Portion of Race effect mediated by anxiety	β for Race, Model 1	R^2 for Race. Model 1	β for Race, Model 2	ΔR^2 for Race, Model 2	Z for difference	Portion of Race effect mediated by anxiety
White vs. African-American (N=1846/1846)	-.48***	23.04%	-.39***	14.1%	9.04***	38.8%	-.40***	15.20%	-.35***	10.90%	6.68***	28.3%
White vs. Latino (N=1047/1047)	-.46***	21.16%	-.37***	12.4%	6.45***	41.4%	-.34***	9.90%	-.27***	6.80%	6.35***	31.3%
White vs. Native American (N=111/111)	-.30***	9.0%	-.21***	4.9%	1.39	45.5%	-.18***	3.1%	-.15*	2.2%	0.86	29.0%

Notes: Model #1 contains only the race variable. For Model #2, race is entered on the second step, following entry of anxiety and the quadratic anxiety variable. For Model #1 in columns where preparation is covaried, prior academic preparation (grades) was entered on the first step and the race variable was entered on the second step. For Model #2 in columns where preparation is covaried, prior academic preparation (grades), anxiety, and quadratic anxiety was entered on the first step and the race variable was entered on the second step. * $p < .05$, *** $p < .0001$.

BEST COPY AVAILABLE

FOOTNOTES

1. Some people appear to be under the impression that stereotype threat occurs only under certain rarified situations, such as when race is specifically cued immediately prior to a test, when the stereotypes are explicitly linked to the task at hand, when the task is very difficult, when the test is clearly diagnostic of ability, and so forth. While experimental studies reported by Steele and others do take pains to ensure that these things occur, they do this for the sake of ensuring that stereotype threat is experienced in a laboratory setting and that they observe a large enough effect size to detect the differences they expect to see.

However, my position is that not all of these elements are necessary to elicit stereotype threat in a naturalistic setting. I don't believe this theory would be the least bit interesting unless it was attempting to describe performance differences that occur in real-world (i.e. natural) settings. Steele's theoretical writings (Steele, 1992, 1997) focus solely on the individual being aware of a negative group stereotype, and of the stereotype becoming situationally self-relevant in a setting that the student is "identified with," (a domain that is self-relevant to the student). These necessary and sufficient conditions are probably present in most school testing situations, from the weekly spelling quiz to the less frequent standardized tests.

2. Three other sections of this test were administered, but were not analyzed in this study as the authors felt they were not representative of most academic tasks: mnemonic memory skills, perceptual comparisons, and three-dimensional visualization. The goal of this study was to examine the stereotype threat hypothesis in the context of tests that students are likely to be familiar with. The stereotype threat hypothesis requires students belonging to a group that stereotypically does poorly on a particular task to know of the stereotype in relation to the task. This would be difficult if a student had never had any experience with a task, such as the three types of tasks excluded from this study (e.g., perceptual comparisons of mosaic patterns). Analyses of these tasks produced effects similar to those reported here, but of smaller magnitude.

3. While this is, admittedly, not the perfect measure of academic preparedness, it is the only variable in this data set that sufficiently captures the construct of interest. Additionally, it does have some evidence for reliability. In other data sets where the same item has been used, agreement is extremely high (average correlations for this item when measuring grades in different content areas are between $\underline{r} = .78$ and $\underline{r} = .82$, depending on the age of the subjects.

4. In order to assert that partial mediation has occurred, it is important to test to see whether a particular regression coefficient (the relationship between race or sex and achievement) changes significantly as the proposed mediators are entered into the equation. In order to test these nested regression models, procedures outlined and discussed in Clogg, Petkova, and Haritou (1995a, 1995b) were used (c.f., Allison, 1995).

BEST COPY AVAILABLE